DOI: 10.1111/1475-6773.13090

RESEARCH ARTICLE

HSR Health Services Research

A comparison of two structured taxonomic strategies in capturing adverse events in U.S. hospitals

John M. Austin PhD[1,2] | Erin M. Kirley BS[1] | Michael A. Rosen PhD[1,2] | Bradford D. Winters PhD, MD, FCCM[2]

[1]Armstrong Institute for Patient Safety and Quality, Johns Hopkins Medicine, Baltimore, Maryland

[2]Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland

Correspondence
John M. Austin, PhD, Armstrong Institute for Patient Safety and Quality, Johns Hopkins Medicine, 750 E. Pratt St, Baltimore, MD 21202.
Email: jausti17@jhmi.edu

Funding information
Agency for Healthcare Research and Quality, Grant/Award Number: HHSP233201500020I

Objective: To compare the Agency for Healthcare Research and Quality's Quality and Safety Review System (QSRS) and the proposed triadic structure for the 11th version of the International Classification of Disease (ICD-11) in their ability to capture adverse events in U.S. hospitals.

Data Sources/Study Setting: One thousand patient admissions between 2014 and 2016 from three general, acute care hospitals located in Maryland and Washington D.C.

Study Design: The admissions chosen for the study were a random sample from all three hospitals.

Data Collection/Extraction Methods: All 1000 admissions were abstracted through QSRS by one set of Certified Coding Specialists and a different set of coders assigned the draft ICD-11 codes. Previously assigned ICD-10-CM codes for 230 of the admissions were also used.

Principal Findings: We found less than 20 percent agreement between QSRS and ICD-11 in identifying the same adverse event. The likelihood of a mismatch between QSRS and ICD-11 was almost twice that of a match. The findings were similar to the agreement found between QSRS and ICD-10-CM in identifying the same adverse event. When coders were provided with a list of potential adverse events, the sensitivity and negative predictive value of ICD-11 improved.

Conclusions: While ICD-11 may offer an efficient way of identifying adverse events, our analysis found that in its draft form, it has a limited ability to capture the same types of events as QSRS. Coders may require additional training on identifying adverse events in the chart if ICD-11 is going to prove its maximum benefit.

KEYWORDS
adverse events, measurement, patient safety

1 | INTRODUCTION

A large number of patients in the United States still experience a preventable harm while receiving care in hospitals.[1] While it is estimated that between 100 000 and 440 000 patients die each year due to preventable harms in U.S. hospitals,[2,3] to date we have not been able to reliably estimate the total number of patients that experience a harm. One challenge in estimating the frequency of harm is we have lacked standard definitions for harms and standardized methods for measuring them. For example, in U.S. hospitals, pressure ulcers may have variable morbidity and consequences for the patient depending on their severity and associated complications and can be tracked as a prevalence rate, an incidence rate, or both rates. Another challenge with estimating harm frequency is we have not had a standard

platform that hospitals and other health care providers can use to capture their harm data in a consistent manner.

The Agency for Healthcare Research and Quality (AHRQ) is currently developing and validating the Quality and Safety Review System (QSRS) to collect comparable patient safety data over time for acute care hospitals using standard definitions and algorithms. QSRS is a web-based software query and reporting system that identifies adverse events based on information abstracted from medical records after discharge.[4] The system is based on the AHRQ Common Formats for Surveillance, which are common definitions and reporting formats designed to help providers uniformly report patient safety events.[5] QSRS is currently a manual system with a maximum of 205 questions answered by a human abstractor and takes an average of 30-60 minutes to abstract the required data elements from the patient's chart for a single hospital stay. The time to abstract depends on the complexity of the patient's case and the organization of the hospital's electronic health record (EHR).[6] Based on two prior rounds of pilot testing of QSRS conducted by the study team in seven acute care hospitals, the observed agreement, or inter-rater reliability (IRR), of two abstractors providing matching answers for individual questions on 700 charts ranged between 74.4 percent and 85.7 percent. The variation in the IRR appears to be linked to the layout of the hospital's EHR and how easily information can be found. The IRR of QSRS is generally better than the IRR that has been measured in previously developed chart abstraction systems such as the Global Trigger Tool.[7] The standardized definitions and algorithms used in QSRS provide the ability to detect a wide variety of harms, and these standard specifications are designed to ensure that a type of adverse event identified is the same across different units and hospitals.[5]

While AHRQ continues to invest in the development of QSRS, other options might exist to complement QSRS in order to more efficiently and completely identify adverse events in EHRs and other databases. For the last 20 years, administrative data, such as billing data, have been a frequently used data source to efficiently capture information on the safety of care.[8-12] Examples of safety measures captured through administrative data include AHRQ's Patient Safety Indicators and 3M's Potentially Preventable Complications.[13,14]

A key coding system used for capturing procedures and patient diagnoses in the hospital setting is the World Health Organization's (WHO) International Classification of Diseases (ICD). The 10th version of ICD was adopted on October 1, 2015, in the United States for coding procedures and diagnoses in the inpatient hospital setting and is referred to as ICD-10-CM. ICD-10-CM greatly expanded the number of codes that had been available with ICD-9-CM—with almost 19 times as many procedure codes and five times as many diagnosis codes.[15] This increase in the number of codes allows for greater specificity of the code, including capture of the etiology, anatomic site, severity, and encounter. The sizable increase in the number of codes that are available, and the improved specificity of those codes, was heralded as an opportunity to capture more detailed information about the quality of the care being delivered in U.S. hospitals.[16] However, despite the expectations for ICD-10-CM, a previous study found "the ability of ICD-10-CM to capture content typically contained in clinical records is not measurably better or worse than that of ICD-9-CM."[17]

The 11th version of ICD (ICD-11) is in development and is expected to be released in 2018.[18] A WHO ICD-11 Quality and Safety Topic Advisory Group (Q&S TAG), which is helping develop the taxonomy for ICD-11, recognized that ICD-11 "presents both a considerable challenge, and also tremendous opportunity, to enhance ICD to better measure quality and safety."[19] ICD-11 as a broader coding system is being designed to allow for cluster coding.[20] The ICD-11 Q&S TAG recognized that cluster coding presents a possible opportunity to more fully capture an adverse event. They have proposed a cluster of three codes for patient safety events—one code that captures the source of harm, one that captures the mode/mechanism of harm, and one that captures the consequence of injury to the patient.[21] For example, if a patient receives an overdose of warfarin that results in an intracerebral hemorrhage, the code cluster would capture the medication (warfarin) as the source of harm, an overdose as the mode/mechanism of harm, and an intracerebral hemorrhage as the consequence of injury to the patient. The three codes describing a patient safety event will be associated with each other by joining the codes together with ampersands.

With the anticipated debut of ICD-11 and its introduction of a new structure for capturing patient safety events in U.S. hospitals, we sought to understand how QSRS and the proposed ICD-11 taxonomic strategies compare to each other in their ability to capture adverse events in U.S. hospitals.

2 | METHODS

2.1 | Setting

To understand how well the two structured taxonomic strategies compare to each other in capturing adverse events in U.S. hospitals, our study used 1000 patient admissions from three general, acute care hospitals located in Maryland and Washington D.C. that had already been processed through QSRS (version 3.1). Eleven Registered Health Information Technician (RHIT) and/or Certified Coding Specialist (CCS) trained coders reviewed the 1000 admissions using QSRS's guided abstraction tool. These abstractors were trained on QSRS by having them attend a 90-minute informational webinar. They then abstracted the same 10 charts and were brought together to discuss their findings and questions from the review of the patient charts. The abstractors were also given the opportunity to ask a clinical expert questions on a weekly call. The primary job of these abstractors was coding for billing, so they were all familiar with medical charts and the locations where clinical documentation could be found. The cases abstracted through QSRS were patient admissions between September 2014 and August 2015 or between January 2016 and June 2016.

2.2 | ICD-11

2.2.1 | How data were collected

Six CCS trained coders were asked to code each of the 1000 admissions that had been processed through QSRS using the draft ICD-11 quality and safety codes that were publicly available in August 2017.[22] This group of coders had not been part of the team that abstracted the charts with QSRS. This set of coders had been trained on ICD-10-CM, and coding medical charts was their primary job function. For training on ICD-11, these coders were provided with background materials to familiarize themselves with ICD-11 and the proposed structure, and then mirroring the training that was provided to the coders who performed the QSRS abstraction, the ICD-11 coders were asked to code 10 training charts. After completing those charts, the coders and authors reviewed the results as a group and discussed improvements that could be made in coding practices to ensure the validity of the codes. The ICD-11 coders also had the opportunity to ask questions as they worked through the charts.

For each of the 1000 admissions, wherever an adverse event was identified by a coder, the coder was asked to identify an ICD-11 code to capture the source of harm and an ICD-11 code to capture the mode/mechanism of harm. Given the almost unlimited consequences of harm that could arise from an adverse event, it was determined that the information that would be gained by coding the consequence would not necessarily be helpful in comparing the two approaches, as almost all possible diagnosis codes would need to be included in the comparison. For example, if a foreign object was left in a patient after surgery, the possible consequence of harm associated with that event could range from an infection, to a blood clot, to nothing at all. As such, ICD-11 coders were instructed that they did not need to identify a code for the consequence of harm.

Through our review of the admissions coded with ICD-11, and in debriefing conversations with the coders, we recognized that the coders were not necessarily familiar with the types of events that would be considered an adverse event, so about three-quarters way through the coding exercise, we provided them with a list of the events that are captured through QSRS in order to familiarize themselves with the types of events they should be looking for in the patient chart. We considered having the coders recode the charts that were done prior to providing them with the list of adverse events, but due to project timeline constraints, we were not able to do this. To understand the potential impact of providing coders with the list of adverse events, we analyzed the coders' overall performance and their performance after providing them with the list of adverse events.

2.2.2 | Data analysis

Once the coding in ICD-11 was complete, we worked with three CCS and/or RHIT trained coders who were part of the QSRS abstraction team and had experience with ICD coding through their daily coding work to develop a crosswalk between each QSRS-identified event and the ICD-11 code, or range of codes, that would be associated with that event. The crosswalk's focus was on identifying the ICD-11 codes for the source of harm and the mode/mechanism of harm, given the almost unlimited consequences of harm that can be associated with an adverse event. The developed crosswalk (Appendix S1) served as a de facto "answer key" between the findings from QSRS and the ICD-11 codes assigned.

The developed crosswalk was used to compare the capture of adverse events by QSRS and ICD-11 in the 1000 chart samples. As some charts used for the analysis had multiple adverse events, our analysis was done at an event level, as opposed to the admission level. For example, if a patient had both a retained foreign object after surgery and a pressure ulcer, we compared the QSRS result and the ICD-11 result for each of the two individual events. This approach provides more granularity to the results and does not necessarily "fail" a chart if one event matches but the other event does not. Additionally, most events are independent of each other (eg, a pressure ulcer has little direct relationship to a retained foreign body) and as such should be compared at the event level.

If one considers the QSRS findings to be a reference standard, perhaps not an unreasonable assumption given the thoroughness of the chart review, both in terms of the time requirement and detailed questions abstractors are asked to answer, one could consider the comparison with ICD-11 in terms of the specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV) of ICD-11 to capture adverse events.

2.3 | ICD-10-CM analysis

To provide context to the comparison between QSRS and ICD-11 and their respective abilities to capture adverse events, we completed a similar comparison between QSRS and ICD-10-CM. The comparison between QSRS and ICD-10-CM was conducted using similar methods to the QSRS/ICD-11 comparison described above. Of the 1000 admissions used in the ICD-11 comparison, 230 were admitted to the hospital between January 2016 and June 2016, which was after ICD-10-CM was introduced in U.S. hospitals. These 230 admissions had already been coded with ICD-10-CM codes by medical billing coders as part of the hospital's regular billing process. We worked with three CCS and/or RHIT trained coders who were experienced with ICD-10-CM coding to develop a crosswalk between each QSRS-identified event and the ICD-10-CM code or codes that would be associated with that event. The developed crosswalk (Appendix S2) served as the "answer key" between the findings from QSRS and the ICD-10-CM codes assigned. The developed crosswalk was used to compare the capture of adverse events by QSRS and ICD-10-CM in the 230 chart samples. Similar to the ICD-11 comparison, the analysis was done at an event level, as opposed to the admission level.

3 | RESULTS

The median age of patients used in the QSRS/ICD-11 comparison was 40 years old, with an interquartile range of 26-67 years old. The

TABLE 1 Details of the admissions in which QSRS identified an event or did not identify an event

	All admissions	Admissions with a QSRS-identified event	Admissions with no QSRS-identified event
Age Median age; [Interquartile range]	40; [26-67]	56; [33-72]	34; [0-57]
Gender	Female 59.6% Male 40.4%	Female 52.3% Male 47.7%	Female 64.6% Male 35.4%
Principal Diagnoses (% of principal diagnoses) Note: diagnosis categories based on ICD-9-CM diagnosis categories; reporting those categories with 10% or more	Complications of pregnancy, childbirth, and the puerperium (23.6%) Live-born infants (16.5%) Injury and poisoning (10.2%)	Injury and poisoning (16.3%) Diseases of the circulatory system (14.3%) Complications of pregnancy, childbirth, and the puerperium (13.0%) Infectious and parasitic diseases (10.5%) Neoplasms (10.3%)	Complications of pregnancy, childbirth, and the puerperium (31.3%) Live-born infants (24.7%)
Principal Procedures (% of principal procedures) Note: procedure categories based on ICD-9-CM procedure categories; reporting those categories with 10% or more	Obstetrical procedures (27.5%) Miscellaneous diagnostic and therapeutic procedures (24.5%) Operations on the cardiovascular system (10.7%)	Operations on the cardiovascular system (18.4%) Miscellaneous diagnostic and therapeutic procedures (17.1%) Obstetrical procedures (13.8%) Operations on the digestive system (13.8%) Operations on the musculoskeletal system (10.0%)	Obstetrical procedures (39.3%) Operations on the integumentary system (30.1%)
Length of Stay Median days; [Interquartile range]	4; [3-10]	9; [4-23]	4; [2-5]
Type of QSRS events identified Note: Due to space considerations, only the largest two categories were reported		Health care-associated infections (182 events) • Surgical site infection (82 events) • Urinary tract infection (63 events) • Others (37 events) Other outcomes of interest (144 events) • Unplanned transfer to higher level of care (54 events) • Respiratory failure (51 events) • Others (39 events)	N/A

median age of patients where QSRS identified an event (56 years old) was older than the median age of patients where ICD-11 identified an event (63 years old; Tables 1 and 2). The overall sample of patients was 59.6 percent female, but only 52.3 percent of events identified by QSRS and 56.9 percent of events identified by ICD-11 occurred in female patients. This overrepresentation of events in male patients may be a function of some female patients being admitted to the hospital for the uncomplicated delivery of a newborn, which is not without the possibility of harm, but those patients are typically younger, healthier, and have a shorter length of stay than other hospitalized patients. The 4-day median length of stay of the admissions in the overall sample was shorter than the median length of stay for admissions where QSRS identified an event (median of 9 days) or ICD-11 identified an event (median of 10 days). The most common principal diagnoses for the overall patient sample were complications of pregnancy, childbirth, and the puerperium; live-born infants; and injury and poisoning. When an event was identified by QSRS or ICD-11, the most common principal diagnoses for these patients were diseases of the circulatory system, infectious and parasitic diseases, and neoplasms.

Of the 1000 admissions reviewed in QSRS and assigned ICD-11 codes, 479 were identified by the QSRS as having at least one adverse event. Across these 479 admissions, a total of 647 adverse events were identified (Table 3). For 244 of the 647 identified events, both QSRS and ICD-11 identified an event. However, those events were not necessarily in agreement based on the developed

TABLE 2 Details of the admissions in which coders coded an event with ICD-11 and those they did not

	All admissions	Admissions for which coders coded an event with ICD-11	Admissions for which coders did not code an event with ICD-11
Age Median age; [Interquartile range]	40; [26.3-67]	63; [44-75]	36; [22-63]
Gender	Female 59.6% Male 40.4%	Female 56.9% Male 43.1%	Female 59.8% Male 40.2%
Principal Diagnoses (% of principal diagnoses) Note: diagnosis categories based on ICD-9-CM diagnosis categories; reporting those categories with 10% or more	Complications of pregnancy, childbirth, and the puerperium (23.6%) Live-born infants (16.5%) Injury and poisoning (10.2%)	Diseases of the circulatory system (21.8%) Injury and poisoning (15.0%) Neoplasms (12.2%)	Complications of pregnancy, childbirth, and the puerperium (26.5%) Live-born infants (19.1%)
Principal procedures (% of principal procedures) Note: procedure categories based on ICD-9-CM procedure categories; reporting those categories with 10% or more	Obstetrical procedures (27.5%) Miscellaneous diagnostic and therapeutic procedures (24.5%) Operations on the cardiovascular system (10.7%)	Operations on the cardiovascular system (25.0%) Operations on the digestive system (16.7%) Miscellaneous diagnostic and therapeutic procedures (15.2%) Operations on the musculoskeletal system (11.4%)	Obstetrical procedures (31.2%) Miscellaneous diagnostic and therapeutic procedures (26.4%)
Length of Stay Median days; [Interquartile range]	4; [3-10]	10; [5-24]	4; [2-8]
Types of ICD-11 events identified		Device (31 events) • Obstruction of device (two events) • Dislodgement, misconnection, or de-attachment (two events) • Infection of device (nine events) • Other specified modes of injury or harm associated with a surgical or other medical devices (eight events) • Mode of injury or harm associated with a surgical or other medical devices, unspecified (10 events) Medication (57 events) • Overdose of substance (one event) • Drug-related injury or harm in context of correct administration or dosage (seven events) • Unspecified appropriateness of dosing or administration (one event) • Medication or substance that is known to be an allergen (two events) • Drug or substance interactions (seven events) • Inappropriate stoppage or discontinuation of drug (two events) • Other specified modes of injury or harm associated with exposure to a drug, medicament, or biological substance (five events) • Mode of injury or harm associated with exposure to a drug, medicament, or biological substance, unspecified (32 events)	

(Continues)

TABLE 2 (Continued)

All admissions	Admissions for which coders coded an event with ICD-11	Admissions for which coders did not code an event with ICD-11
	Procedure (75 events) • Mode of injury or harm associated with a surgical or other medical devices, unspecified (three events) • Cut or puncture (six events) • Other specified modes of injury or harm associated with a surgical or other medical procedures (17 events) • Mode of injury or harm associated with a surgical or other medical procedures, unspecified (49 events)	

TABLE 3 Comparison of admissions with an adverse event captured by QSRS and ICD-11

		Admissions with an adverse event identified in QSRS		
		Yes	No	Total
Admissions with an adverse event identified by ICD-11	Yes	107 admissions (244 events)	24 admissions (24 events)	131 admissions (268 events)
	No	372 admissions (403 events)	497 admissions (0 events)	869 admissions (403 events)
	Total	479 admissions (647 events)	521 admissions (24 events)	1000 admissions (671 events)

crosswalk. For example, in Chart A, QSRS identified an unintended laceration or puncture and so did the ICD-11 coders (ie, ICD-11 coders coded PK40.91 orthopedic procedure associated with injury or harm, open approach; and PK70 cut or puncture, as mode of injury). But in Chart B, QSRS identified a treatment suggestive of transfusion reaction and the ICD-11 coders identified anemia due to acute blood loss during hip arthroplasty. For Chart B, while the events both are related to blood loss and possibly related to each other (ie, blood loss is likely at the root of both needing a transfusion and anemia), they are not the same event per the developed crosswalk. For 403 of the 647 events, QSRS captured an adverse event that was not captured by ICD-11. For 24 events, ICD-11 captured an adverse event that was not captured by QSRS. Using the QSRS findings as the reference standard, the sensitivity of ICD-11 in capturing an event was 22.3 percent, the specificity was 95.3 percent, the PPV was 81.7 percent, and the NPV was 57.2 percent.

The 244 events that were captured by both QSRS and ICD-11 (in the "Yes"/"Yes" quadrant of Table 3) were then classified into one of four categories (Table 4). Forty-four of the 244 events (18 percent) were a match based on the developed crosswalk, indicating that both QSRS and ICD-11 identified the same event type. Eighty-four of 244 events were not a match based on the crosswalk (34 percent). A small number of QSRS-identified events (15 percent) were not clear enough in their definitions to match with specific ICD-11 codes, so it was not possible to determine whether a match existed. And a sizeable number of the events identified in QSRS as being an adverse event (eg, falls, pressure ulcers) did not lend themselves well to the proposed ICD-11 structure of identifying a source and mode/mechanism of harm, as the proposed ICD-11 structure has a finite set of options for the sources of harm (ie, procedures, devices, medications, and others), and for some events, it was difficult to clearly identify a source of the harm. One example of this is a patient falling in the hospital may not be clearly linked to a procedure, device, or medication; the patient may simply have been a frail and elderly patient.

We found that some of QSRS event categories provided greater level of agreement between QSRS and ICD-11 than other event categories (Table 5). "Events in the Surgery/Anesthesia" (54 percent) and "Device" (33 percent) categories had the highest percentage of matching. Events in the "Medications" category had the highest percentage of not matching (78 percent). In general, events in the QSRS "Other Outcomes of Interest" category lacked clear enough descriptions to identify through the crosswalk whether there was a match. Events in the "Birth" category, some of the HAIs (hospital-acquired pneumonia, in particular), falls, pressure ulcers, and DVT/PEs did not lend themselves well to the proposed structure of identifying a specific source and mode/mechanism of harm, but could possibly be captured by existing coding structures, such as a diagnosis code and either a lack of present on admission (POA) indicator or a code that indicates it developed during the hospital stay.

As was previously discussed, we provided the ICD-11 coders with a list of adverse events that QSRS identifies about three-quarters through their 1000 chart samples. For the last 210 admissions where that list was made available, we found the sensitivity of ICD-11 to be

TABLE 4 Agreement in the matching of adverse events captured by both QSRS and ICD-11 based on the developed crosswalk

Matching category	Number of events (% of events)
Events matched based on the crosswalk	44 (18%)
Events did not match based on the crosswalk	84 (34%)
The definition of the QSRS event was not specific enough to assign ICD-11 codes	37 (15%)
The QSRS event does not clearly fit into the ICD-11 source/mode/consequence triadic structure	79 (32%)
Total	244 (100%)

to 39.4 percent, as compared to 22.3 percent in the 1000 total chart samples; the specificity to be 96.2 percent, as compared to 95.4 percent in the full sample; the PPV to be 94.5 percent, as compared to 81.7 percent in the full sample; and the NPV to be 48.4 percent, as compared to 57.2 percent in the full sample (Table S1).

3.1 | Comparison of QSRS and ICD-10

Of the 230 admissions that were admitted to the hospital after October 1, 2015, where ICD-10-CM codes had already been assigned, 203 were identified by QSRS as having at least one adverse event and 27 were identified as not having an adverse event. Across the 203 admissions with at least one adverse event, a total of 371 adverse events were identified (Table S2). For 145 of the 371 identified events, both QSRS and ICD-10-CM identified an event. However, similar to the ICD-11 comparison, those events in the "Yes"/"Yes" quadrant were not necessarily in agreement based on the developed crosswalk. For 226 of the 371 events, QSRS captured an adverse event that was not captured by ICD-10-CM. And for seven events, ICD-10-CM captured an adverse event that was not captured by QSRS.

The 145 events that were captured by both QSRS and ICD-10-CM (in the "Yes"/"Yes" quadrant of Table S2) were then classified into one of four categories (Table S3). Twenty-five of the 145 events (17 percent) were a match based on the developed crosswalk, indicating that both QSRS and ICD-10-CM identified the same event type. Eighty-three of 145 events were not a match based on the crosswalk (57 percent). And for a small number of QSRS-identified events, the QSRS definition was not clear enough to assign specific ICD-10-CM codes (22 percent) or no applicable ICD-10-CM codes could be identified for the event (4 percent), so in these cases, it was not possible to clearly identify if a match existed. If one considers the QSRS findings to be the reference standard, the sensitivity of ICD-10-CM in capturing an event was 29.0 percent, the specificity was 92.5 percent, the PPV was 96.7 percent, and the NPV was 14.8 percent.

4 | CONCLUSIONS

In our study, we found less than 20 percent agreement between QSRS and the proposed ICD-11 triadic structure in capturing the same adverse events in the inpatient hospital setting. Our analysis found the likelihood of a mismatch between QSRS and ICD-11 to be almost twice that of a match and in certain QSRS event categories we found greater levels of agreement with ICD-11, specifically those event categories that align most closely with the sources of harm outlined in the ICD-11 triadic structure (ie, procedure, device, medication, and others). When the ICD-11 coders were provided with a concrete list of adverse events to look for in the last 210 charts, the sensitivity and NPV of ICD-11 to capture adverse events were much better than what was seen with the full sample of admissions.

We are not aware of any other studies that have specifically looked at a comparison between the QSRS structure and the proposed ICD-11 structure in capturing adverse events. However, the ICD-11 Q&S TAG recently completed a small field trial using vignettes to assess the utility of the proposed ICD-11 framework for classifying patient safety events. They found the proposed ICD-11 framework enabled classification of the majority of patient safety events; however, similar to our findings, cases for which it was difficult to link a specific intervention or omission in care to the outcome (eg, fall), classifying the cause and/or mode was problematic.[23]

Other studies that have looked at other administrative data sets, such as ICD-9 and ICD-10, have found mixed results with their ability to identify adverse events.[24-27] Our analysis comparing QSRS and ICD-10-CM found very similar results to the QSRS/ICD-11 comparison in that there was less than 20 percent agreement between QSRS and ICD-10-CM in capturing the same adverse events in the inpatient hospital setting. Although given the smaller sample of admissions that were used for this subanalysis, some caution may need to be used in interpreting these findings.

There were two key limitations to our analysis. First, we recognize that the coders who coded the admissions with ICD-11 were provided with limited training on ICD-11 given its lack of implementation and perhaps the level of training was not sufficient for the task. This is particularly noteworthy as ICD coding is associated with a relatively large knowledge burden, as compared to QSRS, which guides abstractors through questions to identify events. If this exercise is to be repeated, we might suggest finding ways to use coders with greater experience with ICD-11 or identifying a more established training program. Another potential opportunity to improve the ICD-11 coding would be to assess the inter-rater reliability of coders on a sample of charts, something we did not do as part of this initial study. The second limitation was there was a lack of awareness

TABLE 5 Agreement in matching for events captured by both QSRS and ICD-11 by Common Formats Event Category

Common Formats Event Category	No. that matched (% of total)	No. that did not match (% of total)	No. that need greater detail in QSRS/common formats definition (% of total)	No. for which the proposed triadic structure does not fit the event (% of total)	Total
Birth				6 (100%)	6
Blood or blood product		8 (100%)			8
Device	5 (33%)	9 (60%)	1 (7%)		15
Fall				2 (100%)	2
Health care-associated Infections	11 (22%)	16 (32%)		23 (46%)	50
Medications	10 (22%)	36 (78%)			46
Pressure ulcers				32 (100%)	32
Surgery/anesthesia	15 (54%)	11 (39%)	2 (7%)		28
Venous thromboembolism				16 (100%)	16
Other outcomes of interest	3 (7%)	4 (10%)	34 (83%)		41
Total	44 (18%)	84 (34%)	37 (15%)	79 (32%)	244

from the medical coders on what types of events they were looking for in their reviews; that is, what constitutes an adverse event. From debriefings with the coders, it was made clear that they are not necessarily trained to look for adverse events in their review of the chart, unless it is part of coding for reimbursement. This likely is a result of the nature of medical coder education and their primary task of coding medical records for billing purposes as opposed to patient safety and quality purposes. Additionally, medical coders are generally restricted to prescriber documentation for the purposes of developing billing codes for reimbursement and adverse events may only appear in the EHR in other nonprescriber documentation fields (eg, nurses' notes). If ICD coding systems are to be used for capturing adverse events, coders may need greater clinical acumen to accurately capture adverse events, which could require some fundamental changes in their education and/or training.

These two limitations presented a dual test for the coders for the comparison with ICD-11—how well did they understand what a quality and safety event was and how well could they code this event using ICD-11? This duality in their task makes it difficult to assess whether the ICD-11 classification and structure assist in capturing quality and safety events in comparison with the QSRS. We were not able to empirically parse out what piece of the challenge was the lack of understanding of what the coders were looking for and what were the limitations of ICD-11. We did find that the sensitivity, specificity, and PPV of the ICD-11 coding improved once the coders were provided with a list of the types of events that were considered to be adverse events. And while we are careful in interpreting these higher values necessarily as causation between providing the list and better coding performance, one might conclude that the improved performance may illustrate that medical coder training and education in its current form do not focus enough on adverse event recognition and the codes for reporting these types of events and may need to be addressed if ICD-11 is going to prove its maximum benefit.

We identified a number of additional issues that may have contributed to the low level of agreement between the two coding systems. First, QSRS and ICD-11 approach the capture of adverse events in different ways. QSRS uses the Common Formats for Surveillance as a structured, guided algorithm to help coders identify events, while ICD-11 has the proposed framework based on identifying a source, mode, and consequence or harm, but does not guide the coder through the process of identifying events. These different approaches each have their strengths, but do result in differences in the ability to capture specific events; they both may catch equal numbers of events though there may be differences in event types captured. Alternatively, they may catch different numbers of events or possibly a combination of both discrepancies. Given the open-ended approach that is currently used by ICD for identifying adverse events, one recommendation for ICD-11 might be to develop a more structured process for guiding coders through the process of identifying adverse events.

Second, it was difficult for certain QSRS events to fit into the proposed triadic structure (source/mode/consequence cluster) that has been proposed for ICD-11. It appears as if the types of adverse events where the triadic structure is less useful in identifying events are those that do not involve direct action by a health care worker, but where there is some baseline risk, regardless of mitigation strategies in place to prevent them from happening. For example, deep vein thromboembolisms can happen even when appropriate prophylaxis is provided. The usefulness of the triadic structure seems to somewhat map onto latent/active error distinctions. It is much easier to identify a mode or mechanism of harm for an active error as compared to an error with a diffuse or latent origin. What makes errors of a latent origin especially challenging is the general lack of documentation reflecting the omission of care. The structure of using diagnosis codes coupled with a not present on admission indicator, which were previously available in the

United States, may likely be a better approach for capturing events with a diffuse or latent origin.

Finally, some of the QSRS events are a "bad" outcome for the patient (eg, a hemorrhage during childbirth), but are not necessarily always a function of poor care from the hospital. ICD-11 asks a bit more in requiring a mode or mechanism of harm, which does not always exist with some of the events included in QSRS though mechanisms maybe apparent (such as an elevated international normalized ratio and a hemorrhagic complication).

The similarity in our findings between QSRS and ICD-11 and QSRS and ICD-10-CM in their ability to identify the same adverse events highlights the potential challenge that ICD coding, as a broader taxonomy, presents for capturing adverse events in hospitals. Given that we did see improved agreement between ICD-11 and QSRS when coders were given a concrete list of events to look for in the chart, a future study could explore the differences in ICD-10-CM and ICD-11 results if all coders are provided with a list of adverse events to look for in the chart.

The ICD-11 Q&S TAG which is helping develop the framework for ICD-11 recognizes the opportunity they have to help shape ICD into a coding system that better captures adverse events that occur during the patient's hospital stay. And while ICD-11 may offer a more efficient way of identifying adverse events, our analysis has identified that the current open-ended approach that is used in ICD coding may limit the ability of ICD-11 to capture the same types of events as QSRS currently captures, unless substantial education is done with coders on what types of adverse events they are looking for in the chart. This may simply be the natural trade-off between specificity and efficiency. Nevertheless, it may be something that is worth revisiting as both QSRS and ICD-11 further develop.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The funding for this work was provided through an ACTION III contract from the Agency for Healthcare Research and Quality (AHRQ; contract no. HHSP233201500020I). The content of this article is solely the responsibility of the authors and does not represent the official views or recommendations of AHRQ or the Department of Health and Human Services. The staff at AHRQ are aware of, and supportive of, the submission of this manuscript to Health Services Research.

ORCID

John M. Austin http://orcid.org/0000-0002-3061-3540

REFERENCES

1. Kizer KW, Stegun MB. Serious reportable adverse events in health care. In: Henriksen K, Battles JB, Marks ES, et al., eds. *Advances in Patient Safety: From Research to Implementation (Volume 4: Programs, Tools, and Products)*. Rockville, MD: Agency for Healthcare Research and Quality; 2005.
2. Donaldson MS, Corrigan JM, Kohn LT. *To Err is Human: Building a Safer Health System*. Washington DC: National Academies Press; 2000.
3. James JT. A new, evidence-based estimate of patient harms associated with hospital care. *J Patient Saf*. 2013;9(3):122-128.
4. Classen DC, Munier W, Verzier N, et al. Measuring patient safety: the Medicare patient safety monitoring system (past, present, and future). *J Patient Saf*. 2016 [Epub ahead of print]. https://doi.org/10.1097/PTS.0000000000000322
5. Agency for Healthcare Research and Quality. Common Formats. https://pso.ahrq.gov/common. Accessed September 15, 2018.
6. Fong A, Adams K, Samarth A, et al. Assessment of automating safety surveillance from electronic health records: analysis for the quality and safety review system. *J Patient Saf*. 2017 [Epub ahead of print]. https://doi.org/10.1097/PTS.0000000000000402
7. Hibbert PD, Molloy CJ, Hooper TD, et al. The application of the Global Trigger Tool: a systematic review. *Int J Quality Health Care*. 2016;28(6):640-649.
8. Borzecki AM, Kaafarani HM, Utter GH, et al. How valid is the AHRQ patient safety indicator "postoperative respiratory failure"? *J Am Coll Surg*. 2011;212(6):935-945.
9. Cevasco M, Borzecki AM, McClusky DA, et al. Positive predictive value of the AHRQ Patient Safety Indicator "postoperative wound dehiscence". *J Am Coll Surg*. 2011;212(6):962-967.
10. Crews HR, Chamness AR, Terry CL, Helft PR. The use of claims-based data in inpatient public reporting and pay-for-performance programs: is there opportunity for improvement? *J Healthc Qual*. 2016;38(6):379-395.
11. Maass C, Kuske S, Lessing C, Schrappe M. Are administrative data valid when measuring patient safety in hospitals? A comparison of data collection methods using a chart review and administrative data. *Int J Qual Health Care*. 2015;27(4):305-313.
12. Utter GH, Cuny J, Strater A, Silver MR, Hossli S, Romano PS. Variation in academic medical centers' coding practices for postoperative respiratory complications: implications for the AHRQ postoperative respiratory failure Patient Safety Indicator. *Med Care*. 2012;50(9):792-800.
13. Borzecki AM, Chen Q, Mull HJ, et al. Do acute myocardial infarction and heart failure readmissions flagged as potentially preventable by the 3M potentially preventable readmissions software have more process-of-care problems? *Circ Cardiovasc Qual Outcomes*. 2016;9(5):532-541.
14. Hefner JL, Huerta TR, McAlearney AS, Barash B, Latimer T, Moffatt-Bruce SD. Navigating a ship with a broken compass: evaluating standard algorithms to measure patient safety. *J Am Med Inform Assoc*. 2017;24(2):310-315.
15. Centers for Disease Control and Prevention (CDC). International Classification of Diseases, (ICD-10-CM/PCS) Transition – Background. 2015. https://www.cdc.gov/nchs/icd/icd10cm_pcs_background.htm. Accessed September 15, 2018.
16. American Health Information Management Association. ICD-9 CM to ICD-10 CM: Implementation Issues and Challenges. 2003. http://library.ahima.org/doc?oid=59978#.WywVD1VKiM8. Accessed September 15, 2018.
17. Chute CG, Huff SM, Ferguson JA, Walker JM, Halamka JD. There are important reasons for delaying implementation of the new ICD-10 coding system. *Health Aff (Millwood)*. 2012;31(4):836-842.
18. World Health Organization – Health Data Standards and Informatics. ICD-11 Update. 2017. http://who.int/classifications/ICD11January2017Newsletter.pdf. Accessed September 15, 2018.
19. Ghali WA, Pincus HA, Southern DA, et al. ICD-11 for quality and safety: overview of the WHO Quality and Safety Topic Advisory Group. *Int J Qual Health Care*. 2013;25(6):621-625.

20. World Health Organization. ICD-11 Reference Guide Draft 2017-10-9.2017. https://icd.who.int/dev11/Downloads/Download?fileName=refguide.pdf. Accessed September 15, 2018.
21. Southern DA, Pincus HA, Romano PS, et al. Enhanced capture of healthcare-related harms and injuries in the 11th revision of the International Classification of Diseases (ICD-11). *Int J Qual Health Care*. 2016;28(1):136-142.
22. World Health Organization. ICD-11 for Mortality and Morbidity Statistics (May 2018). 2018. https://icd.who.int/browse11/l-m/en. Accessed September 15, 2018.
23. Forster AJ, Bernard B, Drosler SE, et al. A World Health Organization field trial assessing a proposed ICD-11 framework for classifying patient safety events. *Int J Qual Health Care*. 2017;29(4):548-556.
24. Kuklik N, Stausberg J, Jockel KH. Adverse drug events in German hospital routine data: a validation of International Classification of Diseases, 10th revision (ICD-10) diagnostic codes. *PLoS ONE*. 2017;12(11):e0187510.
25. Rajaram R, Barnard C, Bilimoria KY. Concerns about using the patient safety indicator-90 composite in pay-for-performance programs". *JAMA*. 2015;313(9):897-898.
26. Rosen AK, Itani KH, Cevasco M, et al. Validating the patient safety indicators in the Veterans Health Administration: do they accurately identify true safety events? *Med Care*. 2012;50(1):74-85.
27. Winters BD, Bharmal A, Wilson RF, et al. Validity of the agency for health care research and quality patient safety indicators and the Centers for Medicare and Medicaid Hospital-acquired conditions: a systematic review and meta-analysis. *Med Care*. 2016;54(12):1105-1111.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Austin JM, Kirley EM, Rosen MA, Winters BD. A comparison of two structured taxonomic strategies in capturing adverse events in U.S. hospitals. *Health Serv Res*. 2019;54:613–622. https://doi.org/10.1111/1475-6773.13090